Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants?

Lorenzo-Luaces, Lorenzo; Zimmerman, Mark; Cuijpers, Pim

published in
Journal of Affective Disorders
2018

DOI (link to publisher)
10.1016/j.jad.2018.02.066

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Lorenzo-Luaces, L., Zimmerman, M., & Cuijpers, P. (2018). Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants? *Journal of Affective Disorders*, *234*, 8-13. https://doi.org/10.1016/j.jad.2018.02.066

Journal of Affective Disorders 234 (2018) 8–13

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Are studies of psychotherapies for depression more or less generalizable than studies of antidepressants?

Lorenzo Lorenzo-Luaces[a,*], Mark Zimmerman[b], Pim Cuijpers[c]

[a] Indiana University, Department of Psychological and Brain Sciences, 1101 E 10th St, Bloomington, IN 47405, United States
[b] Brown University School of Medicine, Department of Psychiatry and Human Behavior, United States
[c] VU University Amsterdam, Department of Clinical, Neuro, and Developmental Psychology, The Netherlands

ARTICLE INFO

Keywords:
Depression
Psychotherapy
Pharmacotherapy
Generalizability
Exclusion criteria

ABSTRACT

Background: The generalizability of findings from studies exploring the efficacy of psychotherapy and antidepressants has been called into question in part because studies exclude many patients. Despite this, the frequency with which psychotherapy and antidepressant studies use specific inclusion and exclusion criteria has never been compared. We explored the exclusion criteria used in psychotherapy and pharmacotherapy studies from 1995 to 2014.

Method: Systematic literature searches were conducted in PubMed, Medline, PsycINFO, and Embase of published randomized controlled trials (RCTs) of the treatment of major depressive disorder (MDD) in adults with either antidepressants (vs. placebos) or psychotherapy (vs. placebos, treatments as usual, or other controls).

Results: Most psychotherapy (81%) and antidepressant (100%) trials excluded patients with milder symptoms as well as patients with elevated suicidal risk (56–75%), psychotic symptoms (84–88%), or substance misuse (75–81%). Psychotherapy studies were less likely to exclude patients on the basis of brief episode duration (0% vs. 48%) and co-morbid Axis I disorders (6% vs. 27%). However, psychotherapy studies excluded patients with more severe symptoms more frequently (38%) than antidepressant studies (8%).

Conclusions: Overall, psychotherapy studies appear somewhat more inclusive than antidepressant studies. On average, antidepressant studies appear to target patients with more chronic and severe, as well as more purely depressive presentations.

1. Introduction

The generalizability of treatment studies in mental health has been a topic of much debate. More than 20 years ago, Seligman (1995) pointed out that biases in the selection of patient samples in psychotherapy trials, specifically the exclusion of co-morbidities and subclinical presentations, represent a serious threat to the generalizability and applicability of findings from psychotherapy research. Zimmerman et al. (2002) made similar observations regarding the state of the research on medications for depression and provided evidence that most patients seen in outpatient practice were ineligible for a prototypical antidepressant efficacy trial. The issue of the representativeness of participants from treatment trials is of special importance in the study of treatments of major depression because this disorder is characterized by a high degree of heterogeneity in presentation, co-morbid features, and prognosis (Kessler et al., 2016; Lorenzo-Luaces, 2015; Parker, 2005). In an early study addressing this issue, Westen and Morrison (2001) reported that most (68%) patients with depression were excluded from a typical psychotherapy study. Summarily reviewing the literature on psychotherapies for depression, these authors stated that:

> "the prototypical study of treatment for depression excluded patients for suicidality or comorbid substance use disorders. Several studies also excluded patients who had one or more of the following: GAD, panic disorder, antisocial personality disorder, severe obsessional symptoms, schizotypal features, or significant physical problems. Several excluded patients if these comorbid conditions were considered primary but did not define how that determination was made (or report reliability of that determination). The majority of studies required a diagnosis of major depressive disorder for inclusion." (p. 886)

Although exclusion rates appear to be high, as Wiltsey-Stirman et al. (2003) note, many different criteria are used for exclusion and not all are threats to generalizability. For example, a patient may be excluded from an outpatient trial testing the efficacy of a medication for depression if it is deemed that they require a higher level of care (e.g.,

* Corresponding author.
E-mail address: lolorenz@indiana.edu (L. Lorenzo-Luaces).

https://doi.org/10.1016/j.jad.2018.02.066
Received 14 November 2017; Received in revised form 3 February 2018; Accepted 25 February 2018
Available online 27 February 2018

hospitalization) than what is being provided. Such a decision comports with the "real world" delivery of mental health care. It bolsters, and does not threaten, external validity. Thus, it is necessary to consider the specific reasons *why* patients are excluded from treatment trials before assuming that the exclusions are a threat to the external validity of studies.

Recently, Zimmerman and colleagues reviewed the frequency with which inclusion and exclusion criteria are applied in studies exploring the efficacy of antidepressants (Zimmerman et al., 2015, 2016a, 2016b, 2016c). According to these authors, the most commonly used exclusion criteria were: minimum symptom severity (required by 100% of the trials), a diagnosis of a psychotic disorder or current psychotic features (84%), substance abuse or dependence (81%), and significant suicidality (75%). Similar findings were produced by van der Lem et al. (2012), who reported that a minimum symptom severity (80%), a diagnosis of a psychotic disorder or current psychotic features (90%), substance abuse or dependence (85%), and significant suicidality (40%) were frequently used as exclusion criteria for psychotherapy studies.

Huhn et al. (2014) conjectured that psychotherapy trials, by virtue of requiring patients who can actively contribute to therapy have more "well-selected" samples, which are less generalizable than the samples in pharmacotherapy studies. Although a cursory comparison of the studies by van der Lem et al. (2012) and Zimmerman et al. (2015, 2016a, 2016b, 2016c) argues against this conclusion, it is not possible to compare these figures because the authors explored different years of publication and applied different exclusionary criteria. A prior study explored the application of prototypical inclusion criteria in studies for adult depression but without actually studying how often these criteria were employed or comparing them in antidepressant vs. psychotherapy trials (Blanco et al., 2008). To address this gap in the literature, we aimed to compare the frequency with which exclusion criteria are used in trials of antidepressants versus psychotherapy focusing on adults in contemporary trials published between 1995 and 2014. Existing data suggests that psychotherapy studies may be somewhat more generalizable than pharmacotherapy studies for borderline personality (Hoertel et al., 2015), social anxiety (Hoertel et al., 2014), and post-traumatic stress disorder (Franco et al., 2016) though not generalized anxiety disorder (Hoertel et al., 2012). In adolescent depression, the exclusion criteria used in psychotherapy studies tends to exclude fewer adolescents than the exclusion criteria used in pharmacotherapy studies (Blanco et al., 2017). Given these data, we hypothesized that psychotherapy studies would be overall less likely to use psychiatric exclusion criteria than antidepressant studies.

2. Methods

Randomized controlled trials (RCTs) exploring the efficacy of psychotherapy or antidepressants were obtained by referencing recent reviews of the treatment of adult depression (Cuijpers et al., 2008b; Zimmerman et al., 2015) which employed searches in PubMed, Embase, and PsychINFO, and identified individual studies by reviewing meta-analyses and individual journals.

There are systematic differences between antidepressant and psychotherapy studies regarding the type of controls employed in RCTs (Huhn et al., 2014). Virtually all pharmacotherapy studies use a pill placebo whereas the efficacy of psychotherapy is tested with a more diverse mix of controls including: a waiting list (WL), treatment as usual (TAU), pill placebos, and other conditions (e.g., relaxation) that are intended to control for non-specific effects (e.g., attention). Thus, we did not limit our search to the few psychotherapy studies that used a pill placebo and instead included RCTs that used WL, TAU, or other non-therapeutic controls (e.g., psychoeducation). Our search was broadly for psychotherapeutic interventions, irrespective of treatment orientation, in which a therapist delivered material intended to be therapeutic either face to face or by telephone. We excluded guided self-help or internet-based therapy. We included in our search the major therapeutic orientations (e.g., cognitive-behavioral therapy (CBT), behavioral activation (BA), psychodynamic therapy), as well as supportive therapy if it was compared to a control. We required psychotherapy studies to be compared to control conditions so the set of studies would be more comparable to antidepressant-placebo studies (i.e., instead of comparisons of two antidepressants).

Our review was limited to studies of adults with a diagnosis of major depressive disorder not simply major depressive episodes (which could include bipolar depression). We examined acute treatment outcome studies and excluded studies on maintenance treatment and relapse prevention. We excluded trials that focused on co-morbidities, psychiatric or general medical, as they are, by definition less inclusive. For a similar reason, we did not include trials focused on subtypes of depression (e.g., treatment-refractory, chronic, psychotic, atypical, or melancholic). However, we included studies that sampled patients based on demographic features (e.g., low-income, minority). Trials based on inpatients or patients with specific symptoms were also excluded.

The antidepressant efficacy trials reviewed by Zimmerman et al. (2015) were the product of a search through PubMed, Embase, and PsychINFO using the search terms *depression* or *depressive* and *placebo* along with reviews of meta-analyses and the table of contents of 49 journals in which were antidepressant trials are typically published. This process resulted in 170 articles comparing an antidepressant to a placebo. The psychotherapy trials reviewed by Cuijpers et al. (2008b) were the product of a similar search though PubMed, Embase, PsycINFO, and the Cochrane Central Register of Controlled Trials using the search terms *depression* or *depressive* and *various search terms for treatments (e.g., clinical trial, cognitive-behavioral therapy)* along with reviews of meta-analyses. The search process has been described in detail elsewhere and produced a database of comparative psychotherapy outcome studies (Cuijpers et al., 2008b) available in http://www.evidencebasedpsychotherapies.org. The database is updated through 2014 with the search process described by Cuijpers et al., 2008b. The application of our exclusion criteria to this database resulted in 16 studies comparing a psychotherapy to at least one control condition (WL, TAU, or other controls). The list of psychotherapy and antidepressant studies can be found in the Appendix.

One of the authors (MZ) reviewed all of the articles and a second author served as an independent second reviewer. Each reviewer read each article and completed a pre-specified information extraction form listing common psychiatric inclusion and exclusion criteria used in treatment studies. The reviewers compared the results of their data abstraction, and resolved discrepancies. Descriptive analyses summarizing the specific features of the psychotherapy trials are presented. Because a prior study suggested that year of publication was associated with the use of exclusion criteria (Zimmerman et al., 2015), we first compared the year of publication for psychotherapy vs. pharmacotherapy studies using a *t*-test. To compare the differences between antidepressant and psychotherapy studies in the number of studies endorsing specific inclusion criteria we used a chi-square test, or Fisher's exact test when any cell in the 2×2 table was expected to have a frequency lower than 5.

3. Results

Sixteen randomized controlled trials comparing a psychotherapy to a control were analyzed. Most of the psychotherapy studies (81%) explored the efficacy of one of the cognitive-behavioral therapies (CBTs). The other studies explored the efficacy of interpersonal psychotherapy (n = 2) or psychodynamic therapy (n = 1). Most studies used TAU (56%, n = 9) as the control condition. The remaining studies used placebos (n = 4) or waiting lists (n = 2), and a single study used relaxation training as a control condition (Murphy et al., 1995). There were no statistically significant differences ($t(187) = 1.22$, $p = 0.23$) in the average year of publications of psychotherapy (M = 2005.13, SD

Table 1
Frequency of commonly-used psychiatric inclusion and exclusion criteria in randomized controlled trials of antidepressants (n = 170) or psychotherapy (n = 16).

	n	%	*n*	%	*p*
Severity scale score below cutoff	170	100%	13	81%	0.00
Psychotic disorder/current psychotic features	143	84%	14	88%	1.00
Substance abuse/dependence	137	81%	12	75%	0.53
Significant suicidal ideation	128	75%	9	56%	0.13
Episode duration too short	81	48%	2	13%	0.00
Any Axis II disorder	60	35%	3	19%	0.27
Any Axis I disorder	46	27%	1	6%	0.08
History of suicide attempt(s)	35	21%	2	13%	0.74
Episode duration too long	34	20%	2	13%	0.74
Significant homicidal ideation/violence risk	28	16%	0	0%	0.14
Severity scale score above cutoff	14	8%	6	38%	0.00

Note. [a] p value from $\chi 2$ statistic or by Fisher exact test if the expected value in any cell of a 2 × 2 table was less than 5.

= 6.16) vs. antidepressant (M = 2007.04, SD = 6.56) trials.

Table 1 summarizes the clinical/psychiatric exclusion criteria used in the RCTs exploring the efficacy of psychotherapies or antidepressants. Four criteria were used in at least half of the psychotherapy studies: psychotic disorder or current psychotic symptoms (81%), substance abuse or dependence (88%), minimum symptom severity on a depression scale (81%), and significant suicidal ideation (56%). As has been noted of the antidepressant treatment studies, the definition of some exclusion criteria varied between studies. For example, although most of the studies used symptom severity as an exclusion criterion, there was variability in what symptom measure was used to quantify severity and what cut-off was used. Similarly, many studies excluded patients with current or recent substance use disorders but there was substantial variability in whether patients were excluded due heavy substance use, meeting criteria for abuse, or dependence. Virtually all the exclusion criteria we coded were used less frequently in the psychotherapy trials, though not all these differences were statistically significant. The largest observed difference was that around half (48%) of antidepressant studies imposed a minimum episode duration that was longer than the DSM requirement of 2 weeks but none (0%) of the psychotherapy studies explicitly used this exclusion criterion ($p < 0.001$). Two psychotherapy studies required participants to meet the minimum symptom severity in two assessments separated by a two-week period, thus technically imposing minimum depression duration of 4 weeks (i.e., the initial 2 weeks required by the diagnosis of major depression plus the two weeks between the assessment). Re-coding these studies as requiring a minimum duration did not change the result; antidepressants studies were still more likely to use a minimum episode duration criterion (48% vs. 13%, $p < 0.0075$).

Only one psychotherapy efficacy trial (6%) excluded participants on the basis of having *any* other Axis I disorder but this exclusion criterion was present in over a quarter (27%) of the antidepressant trials ($p = 0.08$). Table 2 shows the frequency with which specific disorders were used as exclusion criteria, either explicitly stated or as part of a broader exclusion criterion (e.g., exclusion for any anxiety disorder). As can be seen in the table, some of the differences in the frequency with which studies excluded particular diagnoses were quite large. For example, around half of antidepressant studies excluded subjects on the basis of having posttraumatic stress disorder (PTSD; 44%) or obsessive-compulsive disorder (OCD; 47%) but having these disorders was rarely an exclusion criterion in psychotherapy studies (PTSD: 6%, $p < 0.0001$; OCD: 13%, $p < 0.0001$). Statistical trends in the data suggested that the psychotherapy studies were less likely to exclude based on generalized anxiety, social anxiety, and panic disorder.

Psychotherapy trials were less likely to use as exclusion criterion a minimum score on a symptom severity measure (81% vs. 100%, $p < 0.001$). Furthermore, we observed that even when they used symptom severity to exclude participants, psychotherapy studies were still more inclusive of individuals lower on symptom severity. This observation could be explored quantitatively in the subset of studies that used a common exclusion measure, namely the 17-item version of the Hamilton Rating Scale for depression, which was used in 9 psychotherapy studies (56%) and 105 antidepressants studies (62%). The lowest allowable score on the HRSD-17 for entry into a medication trial was a 14, and the next-most low score was a 15. By way of contrast, the lowest allowable score on the HRSD-17 for entry in a psychotherapy study was a was a 7, and the next lowest score was a 10. Most medication (86%) studies that used the HRSD-17 required that participants have a score of at least 16, which is typically interpreted to reflect depressive symptoms of moderate (Kriston & Wolff, 2011) or mild-to-moderate (Zimmerman et al., 2013) intensity. By contrast, a severity exclusion in the same range was only found in 2 psychotherapy studies (13%; $p < 0.001$), one that required a score of 17 on the HRSD-17 and another that required a score of 20. Six psychotherapy studies used the Beck Depression Inventory (Beck et al., 1996) to exclude on the basis of symptom severity. The range of scores used in these studies was from 10 to 20 and 3 of the 6 studies used scores below 19, which are traditionally interpreted to represent mild depressive symptoms. Taken together, these data would appear to suggest that even when psychotherapy studies used severity exclusions, they allow for the inclusion of individuals with milder depression while medication trials require symptom scores that indicate moderate or severe depression. Although psychotherapy studies were overall less exclusive, and specifically appear more inclusive of milder forms of depression they were more likely to use a maximum score on measures of symptom severity as an exclusion criterion (8% vs. 38%, $p = 0.0029$).

Table 2
Frequency of exclusion of specific Axis I and Axis II disorders in randomized controlled trials of antidepressants (n = 170) or psychotherapy (n = 16).

	n	%	*n*	%	*p*
Drug abuse	130	76%	13	81%	1.000
Alcohol abuse	129	76%	13	81%	0.766
Alcohol dependence	109	64%	13	81%	0.270
Drug dependence	107	63%	13	81%	0.178
Obsessive-compulsive disorder	80	47%	2	13%	0.008
Posttraumatic stress disorder	75	44%	1	6%	0.003
Borderline personality disorder	70	41%	5	31%	0.596
Antisocial personality disorder	68	40%	5	31%	0.598
Bulimia nervosa	66	39%	3	19%	0.012
Anorexia nervosa	66	39%	5	31%	0.603
Panic disorder	65	38%	2	13%	0.055
Schizotypal personality disorder	63	37%	5	31%	0.789
Generalized anxiety disorder	50	29%	1	6%	0.074
Social anxiety disorder	48	28%	1	6%	0.074
Dysthymic disorder	46	27%	2	13%	0.175
Specific phobia	45	26%	1	6%	0.125

Note. [a] p value from $\chi 2$ statistic or by Fisher exact test if the expected value in any cell of a 2 × 2 table was less than 5.

4. Discussion

To our knowledge, this is the first comparison of the specific inclusion and exclusion criteria used in randomized controlled trials of psychotherapies and antidepressants for major depression. On the basis of previous research (Franco et al., 2016; Hoertel et al., 2014, 2015), we hypothesized that research on psychotherapy for adult depression would be more inclusive than research on antidepressants. Most psychotherapy and antidepressant studies excluded patients who were low on symptoms severity, had substance use pathologies, imminent suicidality, or psychotic disorders/symptoms. However, as hypothesized, psychotherapy studies were more inclusive. Specifically, they were less likely to exclude patients on the basis of having brief episodes, lower symptom severity, and specific diagnoses such as OCD, PTSD, and bulimia. Even when they excluded on the basis of symptom severity, the

psychotherapy studies tended to be more inclusive of individuals with milder symptoms. Thus, relative to the research base on psychotherapy, studies of antidepressants appear to select patients with a more severe and chronic, albeit purely depressive, presentation.

Before interpreting the current findings, it is important to note significant limitations of our report. We focused solely on RCTs that included a control condition. Placebos are used almost universally as controls in antidepressant treatment RCTs (Huhn et al., 2014). By contrast, the number of placebo-controlled psychotherapy studies was small (n = 4) and other comparison groups, most often TAU or a waiting list, tended to be used. It may not be appropriate to make comparisons of these two treatment literatures given that the controls differ. Overall, the number of RCTs of psychotherapy was small. Additionally, we focused exclusively on published trials. This introduces the possibility that treatment differences in publication bias somehow relate to differences in the exclusion criteria the trial used. Publication bias relating to efficacy has been documented for both psychotherapy (Driessen et al., 2015) and antidepressant treatments trials (Turner et al., 2008), though it appears to occur with comparable frequency in psychotherapy (24% non-publication rate) and antidepressant trials (31%). Finally, it is worthwhile noting that we focused on the inclusion and exclusion criteria that were listed in the published report. It is possible that antidepressant studies do not use more exclusion criteria than psychotherapy studies but that researchers conducting pharmacological studies are simply more rigorous in their reporting.

A substantial proportion of studies excluded patients due to psychotic symptoms, substance abuse/dependence, and suicidality and the rates did not appear to differ between antidepressant and psychotherapy trials. On the one hand, it might seem appropriate to exclude patients with these clinical features from treatment studies in that they might require a different level of care (e.g., hospitalization) or a treatment that is not being studied (e.g., antipsychotics). On the other hand, the exclusion of these patients from the treatment literature reduces the external validity of the findings from clinical research in ways that have been well -articulated (Morrison et al., 2003; Zimmerman et al., 2002). Excluding these patients from efficacy research represents somewhat of a “Catch-22.” The patients are excluded from research because they are deemed inappropriate, and there is no research speaking to the appropriateness of delivering these treatments to these patients, though they may receive them in the community. There is limited evidence that patients who report psychotic symptoms, problematic substance use, and suicidality experience poorer outcomes in treatment (Kessler et al., 2016). For example, van der Lem et al. (2012) reported that current or past substance use was unrelated to outcomes in a naturalistic sample of depressed patients. However, it would not be unreasonable to assume that outcomes for these patients may vary within medication classes or types of psychotherapy. The fact that these patients are not well represented in the treatment literature thus precludes the identification of specific treatments that may be potentially more efficacious for them or, conversely, treatments that may not help them at all. The fact that these specific exclusion criteria – psychotic symptoms, substance misuse, and suicidality – have been so frequently applied in treatment trials underscores the need for effectiveness research using fewer exclusionary criteria. As well, more attention needs to be paid to developing and testing treatments that target more than one form of psychopathology (e.g., McHugh et al., 2017).

Although, as hypothesized, the overall trend we found was that psychotherapy studies were more inclusive of patients, they were more likely to exclude patients on the basis of having more severe symptoms. This might reflect the enforcement of treatment guidelines (American Psychiatric Association, 2010; Anderson et al., 2008), which were based on a large study suggesting that pharmacotherapy was more effective than CBT for cases of severe depression (Elkin et al., 1995). Individual trials (DeRubeis et al., 2005), study level meta-analyses (Driessen et al., 2010), and individual patient meta-analyses (Weitz et al., 2015) subsequent to this study have failed to replicate this finding and suggest that CBT, and possibly other therapies, can be as effective as medications for more severe depression. Thus, until more evidence emerges that patients with more severe depression have limited success with psychotherapy, or specific kinds of psychotherapy, these individuals should not be excluded from psychotherapy treatment trials. Indeed, these individuals need to be included in treatment trials if any evidence is to emerge suggesting that psychotherapy, or specific kinds of psychotherapy, are less effective than medications for patients with more severe symptoms. Existing data suggest that the combination of medications and psychotherapy are superior to either as a monotherapy (Cuijpers et al., 2010, 2009) so future research should also explore the entry criteria employed in these studies.

One of the most commonly-used criteria was the exclusion of patients with milder symptoms. Although, this was somewhat less likely to happen in psychotherapy studies, most (81%) studies used a minimum severity criterion to exclude patients. Antidepressant studies most often utilized a version of the HRSD to determine study inclusion whereas psychotherapy studies used the HRSD, the BDI, and various other scales. Thus, it may not be appropriate to consider the severity exclusions on equal footing. However, we found some evidence to suggest that even when psychotherapy studies used the HRSD, the cut-off scores tended to be relatively low, thus allowing for the inclusion of patients with mild major depressive disorder. As discussed previously, the targeting of psychotherapy trials to patients with mild to moderate MDD may reflect the belief that psychotherapy is only effective for this patient group (Elkin et al., 1995). The targeting of medication studies to patients with more severe symptoms may reflect a belief that medications are more suitable for cases of more severe depression. Indeed, meta-analyses and analyses of pooled study data casts doubt on the efficacy of antidepressants relative to placebos for less severe depression (Barbui et al., 2011; Fournier et al., 2010; Khan et al., 2002; Kirsch et al., 2008), though these findings have not been universally observed (Rabinowitz et al., 2016), and may not be applicable to more chronic depression or dysthymia (Cuijpers et al., 2008a). It is unclear whether severity is also a moderator of response in psychotherapy vs. control conditions (Driessen et al., 2010; Furukawa et al., 2017). However, the fact that the range of severity is not well-represented in the existing treatment literature is still a cause for concern, limiting even the studies that attempt to explore whether severity is a moderator of treatment outcomes. Nationally representative data suggest that a substantial proportion of individuals who receive prescriptions for antidepressant medications do not meet the full criteria for an anxiety or mood disorder, much less a severe one (Mojtabai and Olfson, 2008; Pagura et al., 2011; Takayanagi et al., 2014) and it is probable that the same is true of patients undergoing psychotherapy. Thus, there is a clear disconnect between the patient populations studied in these trials and the patients who typically receive such treatments, and little in the way of empirical data suggesting how to best treatment milder forms of depressions (Middleton, 2005).

Researchers often perceive a trade-off between internal and external validity (Kazdin, 2016) and associate restrictive exclusion/inclusion criteria with *lower* external validity and *higher* internal validity (Blanco et al., 2008). Most of the studies we analyzed were designed to have high internal validity as they were treatment efficacy trials conducted in academic or academic-medical settings. In these efficacy studies, it may be sensible to employ specific trial entry criteria to maximize detecting treatment effects (i.e., increasing internal validity thereby lowering external validity). The expectation is that subsequent “effectiveness” studies evaluate the generalizability of the findings by relaxing inclusion and exclusion criteria, thus increasing external validity. However, diminished external validity does not guarantee increased internal validity. For example, Hoertel et al. (2013) have argued that some exclusion criteria may lead both to lowered external validity as well as internal validity because the sample under study may underestimate treatment effects. While some of the commonly-used inclusion/exclusion criteria (e.g., mild severity) may increase the likelihood

of finding treatment effects (Fournier et al., 2011), other criteria (e.g., alcohol use disorder) may not (van der Lem et al., 2012). It is also worth noting that concerns have been raised about the transportability of antidepressant (Rutherford et al., 2013) and psychotherapy (Weisz and Gray, 2008) treatments to "real-world" settings. Thus, the inclusion and exclusion criteria are not the only factors bearing on the generalizability of findings from treatment studies.

5. Summary and conclusions

As hypothesized, when differences between the psychotherapy and antidepressant studies emerged, psychotherapy studies seemed less likely to use exclusion criteria. Specifically, these studies were less likely to employ low symptom severity, brief episode duration, and comorbid Axis I pathology as reasons for exclusion. These findings parallel the treatment literature on borderline personality (Hoertel et al., 2015), social anxiety (Hoertel et al., 2014), and post-traumatic stress disorder (Franco et al., 2016) which also suggests that psychotherapy studies might be more representative than pharmacotherapy studies. The findings, however, are somewhat disheartening because most outpatients who are depressed are treated with antidepressant medications instead of psychotherapy (Olfson and Marcus, 2010). Thus, more comparative effectiveness research with less stringent inclusion criteria is needed.

Acknowledgements

The researchers would like to thank Heather Clark, Matt Multach, Emily Walsh, Lia Rosenstein, and Douglas Gazarian for reviewing subsets of antidepressant studies.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2018.02.066.

References

American Psychiatric Association, 2010. American Psychiatric Association Practice Guidelines for the Treatment of Patients With Major Depressive Disorder, 3rd Revision. American Psychiatric Publishing, Washington, D.C. (Retrieved from). ⟨http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf⟩.

Anderson, I., Ferrier, I., Baldwin, R., Cowen, P., Howard, L., Lewis, G., Tylee, A., 2008. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for psychopharmacolog guidelines. J. Psychopharm. 22, 343–396.

Barbui, C., Cipriani, A., Patel, V., Ayuso-Mateos, J.L., van Ommeren, M., 2011. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. Br. J. Psychiatry 198, 11–16.

Beck, A.T., Steer, R.A., Brown, G.K., 1996. Beck Depression Inventory-II. Psychological Corporation, San Antonio, TX.

Blanco, C., Olfson, M., Goodwin, R.D., Ogburn, E., Liebowitz, M.R., Nunes, E.V., Hasin, D.S., 2008. Generalizability of clinical trial results for major depression to community samples: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J. Clin. Psychiatry 69, 1276–1280.

Blanco, C., Hoertel, N., Franco, S., Olfson, M., He, J.P., López, S., Merikangas, K.R., 2017. Generalizability of clinical trial results for adolescent major depressive disorder. Peds e20161701.

Cuijpers, P., van Straten, A., van Oppen, P., Andersson, G., 2008a. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. J. Clin. Psychiatry 69, 1675–1685.

Cuijpers, P., Van Straten, A., Hollon, S.D., Andersson, G., 2010. The contribution of active medication to combined treatments of psychotherapy and pharmacotherapy for adult depression: a meta-analysis. Acta Psychiatr. Scand. 121, 415–423.

Cuijpers, P., van Straten, A., Warmerdam, L., Andersson, G., 2008b. Psychological treatment of depression: a meta-analytic database of randomized studies. BMC Psychiatry 8.

Cuijpers, P., van Straten, A., Warmerdam, L., Andersson, G., 2009. Psychotherapy versus the combination of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. Depression. Anxiety 6, 279–288.

DeRubeis, R.J., Hollon, S.D., Amsterdam, J.D., Shelton, R.C., Young, P.R., Salomon, R.M., Brown, L.L., 2005. Cognitive therapy vs medications in the treatment of moderate to severe depression. Arch. Gen. Psychiatry 62, 409–416.

Driessen, E., Cuijpers, P., Hollon, S.D., Dekker, J.J.M., 2010. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. J. Consult. Clin. Psychol. 78, 668–680.

Driessen, E., Hollon, S.D., Bockting, C.L.H., Cuijpers, P., Turner, E.H., 2015. Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US national institutes of health-funded trials. PLOS ONE 10, e0137864.

Elkin, I., Gibbons, R.D., Shea, M.T., Sotsky, S.M., Watkins, J.T., Pilkonis, P.A., Hedeker, D., 1995. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J. Consult. Clin. Psychol. 63, 841–847.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R.C., Fawcett, J., 2010. Antidepressant drug effects and depression severity. JAMA 303, 47–53.

Franco, S., Hoertel, N., McMahon, K., Wang, S., Rodríguez-Fernández, J.M., Peyre, H., Blanco, C., 2016. Generalizability of pharmacologic and psychotherapy clinical trial results for posttraumatic stress disorder to community samples. J. Psychiatry e975–e981.

Furukawa, T.A., Weitz, E.S., Tanaka, S., Hollon, S.D., Hofmann, S.G., Andersson, G., Hegerl, U., 2017. Initial severity of depression and efficacy of cognitive–behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br. J. Psychiatry 210, 190–196.

Hoertel, N., de Maricourt, P., Katz, J., Doukhan, R., Lavaud, P., Peyre, H., Limosin, F., 2014. Are participants in pharmacological and psychotherapy treatment trials for social anxiety disorder representative of patients in real-life settings? J. Clin. Psychopharm. 34, 697–703.

Hoertel, N., Le Strat, Y., Blanco, C., Lavaud, P., Dubertret, C., 2012. Generalizability of clinical trial results for generalized anxiety disorder to community samples. Depression Anxiety 29, 614–620.

Hoertel, N., Le Strat, Y., Limosin, F., Dubertret, C., Gorwood, P., 2013. Prevalence of subthreshold hypomania and impact on internal validity of RCTs for major depressive disorder: results from a national epidemiological sample. PloS ONE 8, e55448.

Hoertel, N., López, S., Wang, S., González-Pinto, A., Limosin, F., Blanco, C., 2015. Generalizability of pharmacological and psychotherapy clinical trial results for borderline personality disorder to community samples. Personal. Disord. 6, 81–87.

Huhn, M., Tardy, M., Spineli, L.M., Kissling, W., Förstl, H., Pitschel-Walz, G., Leucht, S., 2014. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders. JAMA Psychiatry 71, 706–715.

Kazdin, A.E., 2016. Research Design in Clinical Psychology, Fifth ed. Pearson, Boston.

Kessler, R.C., van Loo, H.M., Wardenaar, K.J., Bossarte, R.M., Brenner, L.A., Ebert, D.D., Zaslavsky, A.M., 2016. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. Epidemiol. Psychiatr. Sci. 26, 22–36.

Khan, A., Leventhal, R.M., Khan, S.R., Brown, W.A., 2002. Severity of depression and response to antidepressants and placebo: an analysis of the food and drug administration database. J. Clin. Psychopharmacol. 22, 40–45.

Kirsch, I., Deacon, B.J., Huedo-Medina, T.B., Scoboria, A., Moore, T.J., Johnson, B.T., 2008. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. PLoS Med. 5, e45.

Kriston, L., von Wolff, A., 2011. Not as golden as standards should be: interpretation of the Hamilton Rating Scale for Depression. J. Affect. Disord. 128, 175–177.

Lorenzo-Luaces, L., 2015. Heterogeneity in the prognosis of major depression: from the common cold to a highly debilitating and recurrent illness. Epidemiol. Psychiatr. Sci. 24, 466–472.

McHugh, R.K., Votaw, V.R., Barlow, D.H., Fitzmaurice, G.M., Greenfield, S.F., Weiss, R.D., 2017. Development of an integrated cognitive behavioral therapy for anxiety and opioid use disorder: study protocol and methods. Cont. Clin. Trials 60, 105–112.

Middleton, H., 2005. NICE guidelines for the management of depression. BMJ 330, 267–268.

Mojtabai, R., Olfson, M., 2008. National patterns in antidepressant treatment by psychiatrists and general medical providers. J. Clin. Psychiatry 69, 1064–1074.

Morrison, K.H., Bradley, R., Westen, D., 2003. The external validity of controlled clinical trials of psychotherapy for depression and anxiety: a naturalistic study. Psychol. Psychother. 76, 109–132.

Murphy, G.E., Carney, R.M., Knesevich, M.A., Wetzel, R.D., Whitworth, P., 1995. Cognitive behavior therapy, relaxation training, and tricyclic antidepressant medication in the treatment of depression. Psychol. Rep. 77, 403–420.

Olfson, M., Marcus, S.C., 2010. National trends in outpatient psychotherapy. Am. J. Psychiatry 1456–1463.

Pagura, J., Katz, L.Y., Mojtabai, R., Druss, B.G., Cox, B., Sareen, J., 2011. Antidepressant Use in the Absence of Common Mental Disorders in the General Population. J. Clin. Psychiatry 72, 494–501.

Parker, G., 2005. Beyond major depression. Psychol. Med. 35, 467–474.

Rabinowitz, J., Werbeloff, N., Mandel, F.S., Menard, F., Marangell, L., Kapur, S., 2016. Initial depression severity and response to antidepressants v. placebo: patient-level data analysis from 34 randomised controlled trials. Br. J. Psychiatry 209, 427–428.

Rutherford, B.R., Cooper, T.M., Persaud, A., Brown, P.J., Sneed, J.R., Roose, S.P., 2013. Less is more in antidepressant clinical trials: a meta-analysis of the effect of visit frequency on treatment response and drop-out. J. Clin. Psychiatry 74, 703–715.

Seligman, M.E.P., 1995. The effectiveness of psychotherapy: the Consumer Reports study. Am. Psychol. 50, 965–974.

Takayanagi, Y., Spira, A.P., Bienvenu, O.J., Hock, R.S., Carras, M.C., Eaton, W.W., Mojtabai, R., 2014. Antidepressant use and lifetime history of mental disorders in a community sample. J. Clin. Psychiatry 40–44.

Turner, E.H., Matthews, A.M., Linardatos, E., Tell, R.A., Rosenthal, R., 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. N. Engl. J. Med. 358, 252–260.

Van der Lem, R., De Wever, W.W., Van der Wee, N.J., Van Veen, T., Cuijpers, P., Zitman, F.G., 2012. The generalizability of psychotherapy efficacy trials in major depressive disorder: an analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. BMC Psychiatry 12, 192.

Weisz, J.R., Gray, J.S., 2008. Evidence-based psychotherapy for children and adolescents: data from the present and a model for the future. Child Adoles. Ment. Health 13, 54–65.

Weitz, E.S., Hollon, S.D., Twisk, J., van Straten, A., Huibers, M.J., David, D., Cristea, I.A., 2015. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: an individual patient data meta-analysis. JAMA Psychiatry 72, 1102–1109.

Westen, D., Morrison, K., 2001. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. J. Consult. Clin. Psychol. 69, 875–899.

Wiltsey-Stirman, S.W., DeRubeis, R.J., Crits-Christoph, P., Brody, P.E., 2003. Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. J. Consult. Clin. Psychol. 71, 963–972.

Zimmerman, M., Clark, H.L., Multach, M.D., Walsh, E., Rosenstein, L.K., Gazarian, D., 2015. Have treatment studies of depression become even less generalizable? A review of the inclusion and exclusion criteria used in placebo-controlled antidepressant efficacy trials published during the past 20 years. Mayo Clin. Proc. 90, 1180–1186.

Zimmerman, M., Clark, H.L., Multach, M.D., Walsh, E., Rosenstein, L.K., Gazarian, D., 2016a. Symptom severity and the generalizability of antidepressant efficacy trials. J. Clin. Psychopharm. 36, 153–156.

Zimmerman, M., Clark, H.L., Multach, M.D., Walsh, E., Rosenstein, L.K., Gazarian, D., 2016b. Variability in the substance use disorder exclusion criterion in antidepressant efficacy trials. J. Affect. Disord. 198, 39–42.

Zimmerman, M., Martinez, J.H., Young, D., Chelminski, I., Dalrymple, K., 2013. Severity classification on the Hamilton depression rating scale. J. Affect. Disord. 150, 384–388.

Zimmerman, M., Mattia, J.I., Posternak, M.A., 2002. Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? Am. J. Psychiatry 159, 469–473.

Zimmerman, M., Multach, M., Walsh, E., Rosenstein, L.K., Gazarian, D., Clark, H.L., 2016c. Problems in the descriptions of the psychiatric inclusion and exclusion criteria in publications of antidepressant efficacy trials: a qualitative review and recommendations for improved clarity. CNS Drugs 30, 185–191.